|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continous |
| Weight of Gold | Continous |
| Distance between two places | Continous |
| Length of a leaf | Continous |
| Dog's weight | Continous |
| Blue Color | Discrete |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |

Q1) Identify the Data type for the Following:

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ordinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Nominal |
| Years of Education | Nominal |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

Ans: When 3 coins tossed, sample space of getting 2 heads and 1 tail is

Sample space = {(H, H, T), (H, T, H), (T, H, H)}

Number of favourable outcomes = 3

Total possible outcomes in 3 tosses = 2^3

= 8

P (Getting 2 heads and 1 tail) =3/8

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3

Ans:

1. P(sum=1) = 0
2. When two dice rolled outcomes with sum <=4 are

{(1,1), (1,2), (2,1), (1,3), (3,1), (2,2)}

Total number of outcomes = 6\*2

= 36

P (sum less than or equal to 4) =1/6

1. Outcomes with sum divisible by 2 and 3 are

S = {(4,2), (2,4), (3,3), (5,1), (1,5), (6,6)}

P (sum is divisible by 2 and 3) = 6 /36

= 1/6

Q5) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

Ans: P (None of the balls drawn is blue) = 5C2/7C2

= 10 /21

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

Ans:

E (Candies for a randomly selected child) = 1\* 0.015 +4\*0.20 +3\*.65+5\*0.005+

6\*0.01+2\*.120

= 3.935

Expected Value = 3.935

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points, Score, Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

Ans:

The Mean values for Points, Score and Weigh are 3.596,3.217 and 17.848

respectively.

The Median values for Points, Score and Weigh are 3.695, 3.325 and 17.71

respectively.

Modes are 3.07 and 3.92 for Points,3.44 for Score and 17.02,18.9 for

Weigh respectively.

The Variance for Points, Score and Weigh are 0.2859, 0.9574 and 3.1932

respectively.

The Standard deviations for Points, Score and Weigh are 0.5347,0.9785

and 1.7869 respectively.

The Range of Points, Score and Weigh are 2.17 ,3.911 and 8.399

respectively.

**Use Q7.csv file**

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Ans: E(X) = (108+110+123+134+145+167+187+ 199)/9 =145.33

Q9) Calculate Skewness, Kurtosis & draw inferences on the following data

Cars speed and distance

Use Q9\_a.csv

SP and Weight(WT)

Use Q9\_b.csv

Q10) Draw inferences about the following boxplot & histogram



Ans:

The given histogram is positively skewed.

The given boxplot is negatively skewed and there are

Outliers.

**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

Ans: 94% Confidence Interval = (198.73, 201.26)

96% Confidence Interval = (198.62, 201.37)

98% Confidence interval = (198.43, 201.56)

**Q12)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median, variance, standard deviation.

Ans: Mean = 41

Median = 40.5

Variance = 25.529

Standard deviation = 5.05

1. What can we say about the student marks?

Ans: Students Marks is positively skewed

Q13) What is the nature of skewness when mean, median of data are equal?

Ans: Perfect symmetry

Q14) What is the nature of skewness when mean > median ?

Ans: positively skewed

Q15) What is the nature of skewness when median > mean?

Ans: Negatively skewed

Q16) What does positive kurtosis value indicates for a data ?

Ans: More peakedness of the data

Q17) What does negative kurtosis value indicates for a data?

Ans: The distribution will be flatter /less peaked

Q18) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

Ans: The distribution is negatively skewed

What is nature of skewness of the data?

Ans: The distribution is negatively skewed

What will be the IQR of the data (approximately)

Ans: IQR = Q3 – Q1

= 18 – 10

= 8

Hence Inter Quartile Range is 8

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

Ans:

For both boxplots the median are same, there are no outliers in 1 and

2, both are symmetric, the range of boxplot 2 is larger compared to boxplot 1.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

Ans:

a) P (MPG > 38) = 0.3475

b) P (MPG < 40) = 0.7293

c) P (20 < MPG <50) = 0.8988

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans:

a) MPG of cars follows Normal Distribution.

b) Adipose tissue follows Normal Distribution and Waist Circumference

does not follows Normal Distribution.

Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

Ans: Z scores of

90% confidence interval = 1.2815

94% confidence interval = 1.5547

60% confidence interval = 0.2533

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

Ans: t scores of

95% confidence interval = 1.7108

96% confidence interval = 1.8280

99% confidence interval = 2.4921

Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

Ans: probability of 18 bulbs randomly selected would have an average

life of no more than 260 days = 0.5